16w

March

RAW SEQUENCE LISTING   DATE: 04/14/2003
PATENT APPLICATION: US/10/007,267A   TIME: 11:04:33

Input Set : **N:\Crf4\Refhold\J007267A.raw**
Output Set: **N:\CRF4\04142003\J007267A.raw**

```
 1 <110> APPLICANT: Gotschlich, Emil C.
 2 <120> TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES,
AND
 3       GENES ENCODING THEM
 4 <130> FILE REFERENCE: 040853-01-5029-02
C--> 5 <140> CURRENT APPLICATION NUMBER: US/10/007,267A
 6 <141> CURRENT FILING DATE: 2003-03-13
 7 <150> PRIOR APPLICATION NUMBER: US 09/333,412
 8 <151> PRIOR FILING DATE: 1999-06-15
 9 <150> PRIOR APPLICATION NUMBER: US 08/878,360
10 <151> PRIOR FILING DATE: 1997-06-18
11 <150> PRIOR APPLICATION NUMBER: US 08/683,426
12 <151> PRIOR FILING DATE: 1996-07-18
13 <150> PRIOR APPLICATION NUMBER: US 08/312,387
14 <151> PRIOR FILING DATE: 1994-09-24
15 <160> NUMBER OF SEQ ID NOS: 13
16 <170> SOFTWARE: PatentIn version 3.2
18 <210> SEQ ID NO: 1
19 <211> LENGTH: 5859
20 <212> TYPE: DNA
21 <213> ORGANISM: Neisseria gonorrheae
22 <220> FEATURE:
23 <221> NAME/KEY: misc_feature
24 <222> LOCATION: (1)..(381)
25 <223> OTHER INFORMATION: glys (glycyl tRNA syntetase beta chain)
26 <220> FEATURE:
27 <221> NAME/KEY: misc_feature
28 <222> LOCATION: (445)..(1491)
29 <223> OTHER INFORMATION: lgtA
30 <220> FEATURE:
31 <221> NAME/KEY: misc_feature
32 <222> LOCATION: (2342)..(3262)
33 <223> OTHER INFORMATION: lgtC
34 <220> FEATURE:
35 <221> NAME/KEY: misc_feature
36 <222> LOCATION: (3322)..(4335)
37 <223> OTHER INFORMATION: lgtD
38 <220> FEATURE:
39 <221> NAME/KEY: misc_feature
40 <222> LOCATION: (4354)..(5196)
41 <223> OTHER INFORMATION: lgtE
42 <400> SEQUENCE: 1
43       ctgcaggccg tcgccgtatt caaacaactg cccgaagccg ccgcgctcgc cgccgccaac    60
44       aaacgcgtgc aaaacctgct gaaaaaagcc gatgccgcgt tgggcgaagt caatgaaagc   120
```

ENTERED

BEST AVAILABLE COPY

RAW SEQUENCE LISTING   DATE: 04/14/2003
PATENT APPLICATION: US/10/007,267A   TIME: 11:04:33

Input Set : N:\Crf4\Refhold\J007267A.raw
Output Set: N:\CRF4\04142003\J007267A.raw

```
45    ctgctgcaac aggacgaaga aaaagccctg tacgctgccg cgcaaggttt gcagccgaaa    180
46    attgccgccg ccgtcgccga aggcaatttc cgaaccgcct tgtccgaact ggcttccgtc    240
47    aagccgcagg ttgatgcctt cttcgacggc gtgatggtga tggcggaaga tgccgccgta    300
48    aaacaaaacc gcctgaacct gctgaaccgc ttggcagagc agatgaacgc ggtggccgac    360
49    atcgcgcttt tgggcgagta accgttgtac agtccaaatg ccgtctgaag ccttcaggcg    420
50    gcatcaaatt atcgggagag taaattgcag cctttagtca gcgtattgat ttgcgcctac    480
51    aacgtagaaa aatattttgc ccaatcatta gccgccgtcg tgaatcagac ttggcgcaac    540
52    ttggatattt tgattgtcga tgacggctcg acagacggca cacttgccat tgccaaggat    600
53    tttcaaaagc gggacagccg tatcaaaatc cttgcacaag ctcaaaattc cggcctgatt    660
54    ccctctttaa acatcgggct ggacgaattg gcaaagtcgg gggggggggg ggggaatat    720
55    attgcgcgca ccgatgccga cgatattgcc tcccccggct ggattgagaa aatcgtgggc    780
56    gagatggaaa aagaccgcag catcattgcg atgggcgcgt ggctggaagt tttgtcggaa    840
57    gaaaaggacg gcaaccggct ggcgcggcac cacaaacacg gcaaaatttg gaaaaagccg    900
58    acccggcacg aagacatcgc cgcctttttc cctttcggca accccataca caacaacacg    960
59    atgattatgc ggcgcagcgt cattgacggc ggtttgcgtt acgacaccga gcgggattgg    1020
60    gcggaagatt accaattttg gtacgatgtc agcaaattgg gcaggctggc ttattatccc    1080
61    gaagccttgg tcaaataccg ccttcacgcc aatcaggttt catccaaaca cagcgtccgc    1140
62    caacacgaaa tcgcgcaagg catccaaaaa accgccagaa acgatttttt gcagtctatg    1200
63    ggttttaaaa cccggttcga cagcctagaa taccgccaaa caaaagcagc ggcgtatgaa    1260
64    ctgccggaga aggatttgcc ggaagaagat tttgaacgcg cccgccggtt tttgtaccaa    1320
65    tgcttcaaac ggacggacac gccgccctcc ggcgcgtggc tggatttcgc ggcagacggc    1380
66    aggatgaggc ggctgtttac cttgaggcaa tacttcggca ttttgtaccg gctgattaaa    1440
67    aaccgccggc aggcgcggtc ggattcggca gggaaagaac aggagattta atgcaaaacc    1500
68    acgttatcag cttggcttcc gccgcagaac gcaggcgca cattgccgca accttcggca    1560
69    gtcgcggcat cccgttccag tttttcgacg cactgatgcc gtctgaaagg ctggaacggg    1620
70    caatggcgga actcgtcccc ggcttgtcgg cgcaccccta tttgagcgga gtggaaaaag    1680
71    cctgctttat gagccacgcc gtattgtggg aacaggcatt ggacgaaggc gtaccgtata    1740
72    tcgccgtatt tgaagatgat gtcttactcg gcgaaggcgc ggagcagttc cttgccgaag    1800
73    atacttggct gcaagaacgc tttgaccccg attccgcctt tgtcgtccgc ttggaaacga    1860
74    tgtttatgca cgtcctgacc tcgccctccg gcgtggcgga ctacggcggg cgcgcctttc    1920
75    cgcttttgga aagcgaacac tgcgggacgg cgggctatat tatttcccga aaggcgatgc    1980
76    gttttttctt ggacaggttt gccgttttgc cgcccgaacg cctgcaccct gtcgatttga    2040
77    tgatgttcgg caaccctgac gacagggaag gaatgccggt ttgccagctc aatcccgcct    2100
78    tgtgcgccca agagctgcat tatgccaagt ttcacgacca aaacagcgca ttgggcagcc    2160
79    tgatcgaaca tgaccgccgc ctgaaccgca aacagcaatg gcgcgattcc cccgccaaca    2220
80    cattcaaaca ccgcctgatc cgcgccttga ccaaaatcgg cagggaaagg gaaaaacgcc    2280
81    ggcaaaggcg cgaacagtta atcggcaaga ttattgtgcc tttccaataa aaggagaaaa    2340
82    gatggacatc gtatttgcgg cagacgacaa ctatgccgcc tacctttgcg ttgcggcaaa    2400
83    aagcgtggaa gcggcccatc ccgatacgga aatcaggttc cacgtcctcg atgccggcat    2460
84    cagtgaggaa aaccgggcgg cggttgccgc caatttgcgg ggggggggta atatccgctt    2520
85    tatagacgta aaccccgaag atttcgccgg cttccccttta aacatcaggc acatttccat    2580
86    tacgacttat gcccgcctga aattgggcga atacattgcc gattgcgaca aagtcctgta    2640
87    tctggatacg gacgtattgg tcagggacgg cctgaagccc ttatgggata ccgatttggg    2700
88    cggtaactgg gtcggcgcgt gcatcgattt gtttgtcgaa aggcaggaag gatacaaaca    2760
89    aaaaatcggt atggcggacg gagaatatta tttcaatgcc ggcgtattgc tgatcaacct    2820
90    gaaaaagtgg cggcggcacg atattttcaa aatgtcctgc gaatgggtgg aacaatacaa    2880
91    ggacgtgatg caatatcagg atcaggacat tttgaacggg ctgtttaaag cggggtgtg    2940
92    ttatgcgaac agccgtttca actttatgcc gaccaattat gcctttatgg cgaacgggtt    3000
93    tgcgtcccgc cataccgacc cgctttacct cgaccgtacc aatacggcga tgcccgtcgc    3060
```

RAW SEQUENCE LISTING   DATE: 04/14/2003
PATENT APPLICATION: **US/10/007,267A**   TIME: 11:04:33

Input Set : **N:\Crf4\Refhold\J007267A.raw**
Output Set: **N:\CRF4\04142003\J007267A.raw**

```
 94       cgtcagccat tattgcggct cggcaaagcc gtggcacagg gactgcaccg tttggggtgc     3120
 95       ggaacgtttc acagagttgg ccggcagcct gacgaccgtt cccgaagaat ggcgcggcaa     3180
 96       acttgccgtc ccgccgacaa agtgtatgct tcaaagatgg cgcaaaaagc tgtctgccag     3240
 97       attcttacgc aagatttatt gacggggcag gccgtctgaa gccttcagac ggcatcggac     3300
 98       gtatcggaaa ggagaaacgg attgcagcct ttagtcagcg tattgatttg cgcctacaac     3360
 99       gcagaaaaat attttgccca atcattggcc gccgtagtgg ggcagacttg gcgcaacttg     3420
100       gatattttga ttgtcgatga cggctcgacg gacggcacgc ccgccattgc ccggcatttc     3480
101       caagaacagg acggcaggat caggataatt tccaatcccc gcaatttggg ctttatcgcc     3540
102       tctttaaaca tcgggctgga cgaattggca aagtcggggg gggggaata tattgcgcgc     3600
103       accgatgccg acgatattgc ctcccccggc tggattgaga aaatcgtggg cgagatggaa     3660
104       aaagaccgca gcatcattgc gatgggcgcg tggttggaag ttttgtcgga agaaaacaat     3720
105       aaaagcgtgc ttgccgccat tgcccgaaac ggcgcaattt gggacaaacc gacccggcat     3780
106       gaagacattg tcgccgtttt ccctttcggc aaccccatac acaacaacac gatgattatg     3840
107       aggcgcagcg tcattgacgg cggtttgcgg ttcgatccag cctatatcca cgccgaagac     3900
108       tataagtttt ggtacgaagc cggcaaactg ggcaggctgg cttattatcc cgaagccttg     3960
109       gtcaaatacc gcttccatca agaccagact tcttccaaat acaacctgca acagcgcagg     4020
110       acggcgtgga aaatcaaaga agaaatcagg gcggggtatt ggaaggcggc aggcatagcc     4080
111       gtcggggcgg actgcctgaa ttacgggctt ttgaaatcaa cggcatatgc gttgtacgaa     4140
112       aaagccttgt ccggacagga tatcggatgc ctccgcctgt tcctgtacga atatttcttg     4200
113       tcgttggaaa agtattcttt gaccgatttg ctggatttct tgacagaccg cgtgatgagg     4260
114       aagctgtttg ccgcaccgca atataggaaa atcctgaaaa aaatgttacg cccttggaaa     4320
115       taccgcagct attgaaaccg aacaggataa atcatgcaaa accacgttat cagcttggct     4380
116       tccgccgcag agcgcagggc gcacattgcc gataccttcg gcagtcgcgg catcccgttc     4440
117       cagtttttcg acgcactgat gccgtctgaa aggctggaac aggcgatggc ggaactcgtc     4500
118       cccggcttgt cggcgcaccc ctatttgagc ggagtggaaa aagcctgctt tatgagccac     4560
119       gccgtattgt gggaacaggc gttggatgaa ggtctgccgt atatcgccgt atttgaggac     4620
120       gacgttttac tcggcgaagg cgcggagcag ttccttgccg aagatacttg gttggaagag     4680
121       cgttttgaca aggattccgc ctttatcgtc cgtttggaaa cgatgtttgc gaaagttatt     4740
122       gtcagaccgg ataaagtcct gaattatgaa aaccggtcat ttcctttgct ggagagcgaa     4800
123       cattgtggga cggctggcta tatcatttcg cgtgaggcga tgcggttttt cttggacagg     4860
124       tttgccgttt tgccgccaga gcggattaaa gcggtagatt tgatgatgtt tacttatttc     4920
125       tttgataagg aggggatgcc tgtttatcag gttagtcccg ccttatgtac ccaagaattg     4980
126       cattatgcca agtttctcag tcaaaacagt atgttgggta gcgatttgga aaaagatagg     5040
127       gaacaaggaa gaagacaccg ccgttcgttg aaggtgatgt ttgacttgaa gcgtgctttg     5100
128       ggtaaattcg gtaggaaaa gaagaaaaga atggagcgtc aaaggcaggc ggagcttgag     5160
129       aaagtttacg gcaggcgggt catattgttc aaatagtttg tgtaaaatat aggggattaa     5220
130       aatcagaaat ggacacactg tcattcccgc gcaggcggga atctaggtct ttaaacttcg     5280
131       gttttttccg ataaattctt gccgcattaa aattccagat tcccgctttc gcggggatga     5340
132       cggcgggggg attgttgctt tttcggataa aatcccgtgt tttttcatct gctaggtaaa     5400
133       atcgccccaa agcgtctgca tcgcggcgat ggcggcgagt ggggcggttt ctgtgcgtaa     5460
134       aatccgtttt ccgagtgtaa ccgcctgaaa gccggcttca aatgcctgtt gttcttcctg     5520
135       ttctgtccag ccgccttcgg gcccgaccat aaagacgatt gcgccggacg ggtggcggat     5580
136       gtcgccgagt ttgcaggcgc ggttgatgct cataatcagc ttggtgtttt cagacggcat     5640
137       tttgtcgagt gcttcacggt agccgatgat gggcagtacg gggggaacgg tgttcctgcc     5700
138       gctttgttcg cacgcggaga tgacgatttc ctgccagcgt gcgaggcgtt ggcggcgcg      5760
139       ttctccgtcg aggcggacga tgcagcgttc gctgatgacg ggctgtatgg cggttacgcc     5820
140       gagttcgacg cttttttgca gggtgaaatc catgcgatc                            5859
142 <210> SEQ ID NO: 2
143 <211> LENGTH: 126
```

RAW SEQUENCE LISTING   DATE: 04/14/2003
PATENT APPLICATION: US/10/007,267A   TIME: 11:04:33

Input Set : N:\Crf4\Refhold\J007267A.raw
Output Set: N:\CRF4\04142003\J007267A.raw

```
144 <212> TYPE: PRT
145 <213> ORGANISM: Neisseria gonorrheae
146 <400> SEQUENCE: 2
147     Leu Gln Ala Val Ala Val Phe Lys Gln Leu Pro Glu Ala Ala Ala Leu
148     1               5                   10                  15
149     Ala Ala Ala Asn Lys Arg Val Gln Asn Leu Leu Lys Lys Ala Asp Ala
150                 20                  25                  30
151     Ala Leu Gly Glu Val Asn Glu Ser Leu Leu Gln Gln Asp Glu Glu Lys
152             35                  40                  45
153     Ala Leu Tyr Ala Ala Ala Gln Gly Leu Gln Pro Lys Ile Ala Ala Ala
154         50                  55                  60
155     Val Ala Glu Gly Asn Phe Arg Thr Ala Leu Ser Glu Leu Ala Ser Val
156     65                  70                  75                  80
157     Lys Pro Gln Val Asp Ala Phe Phe Asp Gly Val Met Val Met Ala Glu
158                 85                  90                  95
159     Asp Ala Ala Val Lys Gln Asn Arg Leu Asn Leu Leu Asn Arg Leu Ala
160                 100                 105                 110
161     Glu Gln Met Asn Ala Val Ala Asp Ile Ala Leu Leu Gly Glu
162             115                 120                 125
164 <210> SEQ ID NO: 3
165 <211> LENGTH: 348
166 <212> TYPE: PRT
167 <213> ORGANISM: Neisseria gonorrheae
168 <400> SEQUENCE: 3
169     Met Gln Pro Leu Val Ser Val Leu Ile Cys Ala Tyr Asn Val Glu Lys
170     1               5                   10                  15
171     Tyr Phe Ala Gln Ser Leu Ala Ala Val Val Asn Gln Thr Trp Arg Asn
172                 20                  25                  30
173     Leu Asp Ile Leu Ile Val Asp Asp Gly Ser Thr Asp Gly Thr Leu Ala
174             35                  40                  45
175     Ile Ala Lys Asp Phe Gln Lys Arg Asp Ser Arg Ile Lys Ile Leu Ala
176         50                  55                  60
177     Gln Ala Gln Asn Ser Gly Leu Ile Pro Ser Leu Asn Ile Gly Leu Asp
178     65                  70                  75                  80
179     Glu Leu Ala Lys Ser Gly Gly Gly Gly Gly Glu Tyr Ile Ala Arg Thr
180                 85                  90                  95
181     Asp Ala Asp Asp Ile Ala Ser Pro Gly Trp Ile Glu Lys Ile Val Gly
182                 100                 105                 110
183     Glu Met Glu Lys Asp Arg Ser Ile Ile Ala Met Gly Ala Trp Leu Glu
184             115                 120                 125
185     Val Leu Ser Glu Glu Lys Asp Gly Asn Arg Leu Ala Arg His His Lys
186         130                 135                 140
187     His Gly Lys Ile Trp Lys Lys Pro Thr Arg His Glu Asp Ile Ala Ala
188     145                 150                 155                 160
189     Phe Phe Pro Phe Gly Asn Pro Ile His Asn Asn Thr Met Ile Met Arg
190                 165                 170                 175
191     Arg Ser Val Ile Asp Gly Gly Leu Arg Tyr Asp Thr Glu Arg Asp Trp
192                 180                 185                 190
193     Ala Glu Asp Tyr Gln Phe Trp Tyr Asp Val Ser Lys Leu Gly Arg Leu
```

```
194              195            200            205
195        Ala Tyr Tyr Pro Glu Ala Leu Val Lys Tyr Arg Leu His Ala Asn Gln
196            210            215            220
197        Val Ser Ser Lys His Ser Val Arg Gln His Glu Ile Ala Gln Gly Ile
198        225            230            235            240
199        Gln Lys Thr Ala Arg Asn Asp Phe Leu Gln Ser Met Gly Phe Lys Thr
200            245            250            255
201        Arg Phe Asp Ser Leu Glu Tyr Arg Gln Thr Lys Ala Ala Ala Tyr Glu
202            260            265            270
203        Leu Pro Glu Lys Asp Leu Pro Glu Glu Asp Phe Glu Arg Ala Arg Arg
204            275            280            285
205        Phe Leu Tyr Gln Cys Phe Lys Arg Thr Asp Thr Pro Pro Ser Gly Ala
206            290            295            300
207        Trp Leu Asp Phe Ala Ala Asp Gly Arg Met Arg Arg Leu Phe Thr Leu
208        305            310            315            320
209        Arg Gln Tyr Phe Gly Ile Leu Tyr Arg Leu Ile Lys Asn Arg Arg Gln
210            325            330            335
211        Ala Arg Ser Asp Ser Ala Gly Lys Glu Gln Glu Ile
212            340            345
214 <210> SEQ ID NO: 4
215 <211> LENGTH: 306
216 <212> TYPE: PRT
217 <213> ORGANISM: Neisseria gonorrheae
218 <400> SEQUENCE: 4
219        Met Asp Ile Val Phe Ala Ala Asp Asp Asn Tyr Ala Ala Tyr Leu Cys
220        1              5              10             15
221        Val Ala Ala Lys Ser Val Glu Ala Ala His Pro Asp Thr Glu Ile Arg
222            20             25             30
223        Phe His Val Leu Asp Ala Gly Ile Ser Glu Glu Asn Arg Ala Ala Val
224            35             40             45
225        Ala Ala Asn Leu Arg Gly Gly Gly Asn Ile Arg Phe Ile Asp Val Asn
226            50             55             60
227        Pro Glu Asp Phe Ala Gly Phe Pro Leu Asn Ile Arg His Ile Ser Ile
228        65             70             75             80
229        Thr Thr Tyr Ala Arg Leu Lys Leu Gly Glu Tyr Ile Ala Asp Cys Asp
230            85             90             95
231        Lys Val Leu Tyr Leu Asp Thr Asp Val Leu Val Arg Asp Gly Leu Lys
232            100            105            110
233        Pro Leu Trp Asp Thr Asp Leu Gly Gly Asn Trp Val Gly Ala Cys Ile
234            115            120            125
235        Asp Leu Phe Val Glu Arg Gln Glu Gly Tyr Lys Gln Lys Ile Gly Met
236            130            135            140
237        Ala Asp Gly Glu Tyr Tyr Phe Asn Ala Gly Val Leu Leu Ile Asn Leu
238        145            150            155            160
239        Lys Lys Trp Arg Arg His Asp Ile Phe Lys Met Ser Cys Glu Trp Val
240            165            170            175
241        Glu Gln Tyr Lys Asp Val Met Gln Tyr Gln Asp Gln Asp Ile Leu Asn
242            180            185            190
243        Gly Leu Phe Lys Gly Gly Val Cys Tyr Ala Asn Ser Arg Phe Asn Phe
```